(51) International Patent Classification 6 :	(11) International Publication Number: WO 95/22625
C12Q 1/68 **A1**	(43) International Publication Date: 24 August 1995 (24.08.95)

(21) International Application Number: PCT/US95/02126

(22) International Filing Date: 17 February 1995 (17.02.95)

(30) Priority Data:
08/198,431   17 February 1994 (17.02.94)   US

(60) Parent Application or Grant
  (63) Related by Continuation
    US   08/198,431 (CIP)
    Filed on   17 February 1994 (17.02.94)

(71) Applicant (for all designated States except US): AFFYMAX
    TECHNOLOGIES N.V. [NL/NL]; De Ruyderkade 62,
    Curaçao (AN).

(72) Inventors; and
(75) Inventors/Applicants (for US only): STEMMER, Willem,
    P., C. [NL/US]; 108 Katy Court, Los Gatos, CA 95030
    (US). CRAMERI, Andreas [CH/US]; 531 Drucilla Drive,
    Mountain View, CA 94040 (US).

(74) Agents: DUNN, Tracy, J. et al.; Townsend and Townsend
    Khourie and Crew, One Market Plaza, 20th floor, Steuart
    Street Tower, San Francisco, CA 94105 (US).

(81) Designated States: AM, AT, AU, BB, BG, BR, BY, CA, CH,
    CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, JP, KE, KG,
    KP, KR, KZ, LK, LR, LT, LU, LV, MD, MG, MN, MW,
    MX, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SI, SK, TJ,
    TT, UA, UG, US, UZ, VN, European patent (AT, BE, CH,
    DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE),
    OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR,
    NE, SN, TD, TG), ARIPO patent (KE, MW, SD, SZ, UG).

Published
    With international search report.

(54) Title: DNA MUTAGENESIS BY RANDOM FRAGMENTATION AND REASSEMBLY

(57) Abstract

A method for DNA reassembly after random fragmentation, and its application to mutagenesis of nucleic acid sequences by *in vitro* or *in vivo* recombination is described. In particular, a method for the production of nucleic acid fragments or polynucleotides encoding mutant proteins is described. The present invention also relates to a method of repeated cycles of mutagenesis, shuffling and selection which allow for the directed molecular evolution *in vitro* or *in vivo* of proteins.

DNA mutagenesis by random fragmentation and reassembly.

BACKGROUND OF THE INVENTION

Field of the Invention

The present invention relates to a method for the production
of polynucleotides conferring a desired phenotype and/or encoding
10   a protein having an advantageous predetermined property which is
selectable. In an aspect, the method is used for generating and
selecting nucleic acid fragments encoding mutant proteins.

Description of the Related Art

15   The complexity of an active sequence of a biological
macromolecule, e.g. proteins, DNA etc., has been called its
information content ("IC"; 5-9). The information content of a
protein has been defined as the resistance of the active protein
to amino acid sequence variation, calculated from the minimum
20   number of invariable amino acids (bits) required to describe a
family of related sequences with the same function (9, 10).
Proteins that are sensitive to random mutagenesis have a high
information content. In 1974, when this definition was coined,
protein diversity existed only as taxonomic diversity.

25   Molecular biology developments such as molecular libraries
have allowed the identification of a much larger number of
variable bases, and even to select functional sequences from
random libraries. Most residues can be varied, although
typically not all at the same time, depending on compensating
30   changes in the context. Thus a 100 amino acid protein can
contain only 2,000 different mutations, but 20^{100} possible
combinations of mutations.

Information density is the Information Content/unit length
of a sequence. Active sites of enzymes tend to have a high
35   information density. By contrast, flexible linkers in enzymes
have a low information density (8).

Current methods in widespread use for creating mutant
proteins in a library format are error-prone polymerase chain
reaction (11, 12, 19) and cassette mutagenesis (8, 20, 21, 22,
40   40, 41, 42), in which the specific region to be optimized is

replaced with a synthetically mutagenized oligonucleotide.   In
both cases, a 'mutant cloud' (4) is generated around certain
sites in the original sequence.

Error-prone PCR uses low-fidelity polymerization conditions
5   to introduce a low level of point mutations randomly over a long
sequence.   Error prone PCR can be used to mutagenize a mixture of
fragments of unknown sequence.   However, computer simulations
have suggested that point mutagenesis alone may often be too
gradual to allow the block changes that are required for
10 continued sequence evolution.   The published error-prone PCR
protocols do not allow amplification of DNA fragments greater
than 0.5 to 1.0 kb, limiting their practical application.
Further, repeated cycles of error-prone PCR lead to an
accumulation of neutral mutations, which, for example, may make
15 a protein immunogenic.

In oligonucleotide-directed mutagenesis, a short sequence is
replaced with a synthetically mutagenized oligonucleotide. This
approach does not generate combinations of distant mutations and
is thus not combinatorial. The limited library size relative to
20 the vast sequence length means that many rounds of selection are
unavoidable for protein optimization. Mutagenesis with synthetic
oligonucleotides requires sequencing of individual clones after
each selection round followed by grouping into families,
arbitrarily choosing a single family, and reducing it to a
25 consensus motif, which is resynthesized and reinserted into a
single gene followed by additional selection.   This process
constitutes a statistical bottleneck, it is labor intensive and
not practical for many rounds of mutagenesis.

Error-prone PCR and oligonucleotide-directed mutagenesis are
30 thus useful for single cycles of sequence fine tuning but rapidly
become limiting when applied for multiple cycles.

Error-prone PCR can be used to mutagenize a mixture of
fragments of unknown sequence (11, 12).   However, the published
error-prone PCR protocols (11, 12) suffer from a low processivity
35 of the polymerase.   Therefore, the protocol is unable to result
in the random mutagenesis of an average-sized gene.   This
inability limits the practical application of error-prone PCR.

Another serious limitation of error-prone PCR is that the

rate of down-mutations grows with the information content of the
sequence.   At a certain information content, library size, and
mutagenesis rate, the balance of down-mutations to up-mutations
will statistically prevent the selection of further improvements
5 (statistical ceiling).

Finally, repeated cycles of error-prone PCR will also lead
to the accumulation of neutral mutations, which can affect, for
example, immunogenicity but not binding affinity.

Thus error-prone PCR was found to be too gradual to allow
10 the block changes that are required for continued sequence
evolution (1, 2).

In cassette mutagenesis, a sequence block of a single
template is typically replaced by a (partially) randomized
sequence. Therefore, the maximum information content that can be
15 obtained is statistically limited by the number of random
sequences (i.e., library size). This constitutes a statistical
bottleneck, eliminating other sequence families which are not
currently best, but which may have greater long term potential.

Further, mutagenesis with synthetic oligonucleotides
20 requires sequencing of individual clones after each selection
round (20). Therefore, this approach is tedious and is not
practical for many rounds of mutagenesis.

Error-prone PCR and cassette mutagenesis are thus best
suited and have been widely used for fine-tuning areas of
25 comparatively low information content. One apparent exception is
the selection of an RNA ligase ribozyme from a random library
using many rounds of amplification by error-prone PCR and
selection (13).

It is becoming increasingly clear that the tools for the
30 design of recombinant linear biological sequences such as
protein, RNA and DNA are not as powerful as the tools nature has
developed. Finding better and better mutants depends on
searching more and more sequences within larger and larger
libraries, and increasing numbers of cycles of mutagenic
35 amplification and selection are necessary. However as discussed
above, the existing mutagenesis methods that are in widespread
use have distinct limitations when used for repeated cycles.

Evolution of most organisms occurs by natural selection and

sexual reproduction.   Sexual reproduction ensures mixing and combining of the genes of the offspring of the selected individuals.   During meiosis, homologous chromosomes from the parents line up with one another and cross-over part way along

5 their length, thus swapping genetic material.   Such swapping or shuffling of the DNA allows organisms to evolve more rapidly (1, 2).   In sexual recombination, because the inserted sequences were of proven utility in a homologous environment, the inserted sequences are likely to still have substantial information

10 content once they are inserted into the new sequence.

Marton et al.,(27) describes the use of PCR _in vitro_ to monitor recombination in a plasmid having directly repeated sequences.   Marton et al. discloses that recombination will occur during PCR as a result of breaking or nicking of the DNA.   This

15 will give rise to recombinant molecules.   Meyerhans et al. (23) also disclose the existence of DNA recombination during _in vitro_ PCR.

The term Applied Molecular Evolution ("AME") means the application of an evolutionary design algorithm to a specific,

20 useful goal.   While many different library formats for AME have been reported for polynucleotides (3, 11-14), peptides and proteins (phage (15-17), lacI (18) and polysomes, in none of these formats has recombination by random cross-overs been used to deliberately create a combinatorial library.

25 Theoretically there are 2,000 different single mutants of a 100 amino acid protein.   A protein of 100 amino acids has 20^{100} possible combinations of mutations, a number which is too large to exhaustively explore by conventional methods.   It would be advantageous to develop a system which would allow the

30 generation and screening of all of these possible combination mutations.

Winter and coworkers (43,44) have utilized an _in vivo_ site specific recombination system to combine light chain antibody genes with heavy chain antibody genes for expression in a phage

35 system.   However, their system relies on specific sites of recombination and thus is limited.   Hayashi et al. (48) report simultaneous mutagenesis of antibody CDR regions in single chain antibodies (scFv) by overlap extension and PCR.

Caren et al. (45) describe a method for generating a large population of multiple mutants using random _in vivo_ recombination. However, their method requires the recombination of two different libraries of plasmids, each library having a different selectable marker. Thus the method is limited to a finite number of recombinations equal to the number of selectable markers existing, and produces a concomitant linear increase in the number of marker genes linked to the selected sequence(s).

Calogero et al. (46) and Galizzi et al. (47) report that _in vivo_ recombination between two homologous but truncated insect-toxin genes on a plasmid can produce a hybrid gene. Radman et al. (49) report _in vivo_ recombination of substantially mismatched DNA sequences in a host cell having defective mismatch repair enzymes, resulting in hybrid molecule formation.

It would be advantageous to develop a method for the production of mutant proteins which method allowed for the development of large libraries of mutant nucleic acid sequences which were easily searched. The invention described herein is directed to the use of repeated cycles of point mutagenesis, nucleic acid shuffling and selection which allow for the directed molecular evolution _in vitro_ of highly complex linear sequences, such as proteins through random recombination.

Accordingly, it would be advantageous to develop a method which allows for the production of large libraries of mutant DNA, RNA or proteins and the selection of particular mutants for a desired goal. The invention described herein is directed to the use of repeated cycles of mutagenesis, _in vivo_ recombination and selection which allow for the directed molecular evolution _in vivo_ of highly complex linear sequences, such as DNA, RNA or proteins through recombination.

Further advantages of the present invention will become apparent from the following description of the invention with reference to the attached drawings.

SUMMARY OF THE INVENTION

The present invention is directed to a method for generating a selected polynucleotide sequence or population of selected polynucleotide sequences, typically in the form of amplified

and/or cloned polynucleotides, whereby the selected polynucleotide sequence(s) possess a desired phenotypic characteristic (e.g., encode a polypeptide, promote transcription of linked polynucleotides, bind a protein, and the like) which

5 can be selected for. One method of identifying polypeptides that possess a desired structure or functional property, such as binding to a predetermined biological macromolecule (e.g., a receptor), involves the screening of a large library of polypeptides for individual library members which possess the

10 desired structure or functional property conferred by the amino acid sequence of the polypeptide.

The present invention provides a method for generating libraries of displayed polypeptides or displayed antibodies suitable for affinity interaction screening or phenotypic

15 screening. The method comprises (1) obtaining a first plurality of selected library members comprising a displayed polypeptide or displayed antibody and an associated polynucleotide encoding said displayed polypeptide or displayed antibody, and obtaining said associated polynucleotides or copies thereof wherein said

20 associated polynucleotides comprise a region of substantially identical sequence, optionally introducing mutations into said polynucleotides or copies, and (2) pooling and fragmenting, typically randomly, said associated polynucleotides or copies to form fragments thereof under conditions suitable for PCR

25 amplification, performing PCR amplification and optionally mutagenesis, and thereby homologously recombining said fragments to form a shuffled pool of recombined polynucleotides, whereby a substantial fraction (e.g., greater than 10 percent) of the recombined polynucleotides of said shuffled pool are not present

30 in the first plurality of selected library members, said shuffled pool composing a library of displayed polypeptides or displayed antibodies suitable for affinity interaction screening. Optionally, the method comprises the additional step of screening the library members of the shuffled pool to identify individual

35 shuffled library members having the ability to bind or otherwise interact (e.g., such as catalytic antibodies) with a predetermined macromolecule, such as for example a proteinaceous receptor, peptide, oligosaccharide, virion, or other

predetermined compound or structure. The displayed polypeptides, antibodies, peptidomimetic antibodies, and variable region sequences that are identified from such libraries can be used for therapeutic, diagnostic, research, and related purposes (e.g., catalysts, solutes for increasing osmolarity of an aqueous solution, and the like), and/or can be subjected to one or more additional cycles of shuffling and/or affinity selection. The method can be modified such that the step of selecting is for a phenotypic characteristic other than binding affinity for a predetermined molecule (e.g., for catalytic activity, stability, oxidation resistance, drug resistance, or detectable phenotype conferred on a host cell).

In one embodiment, the first plurality of selected library members is fragmented and homologously recombined by PCR in vitro.

In one embodiment, the first plurality of selected library members is fragmented in vitro, the resultant fragments transferred into a host cell or organism and homologously recombined to form shuffled library members in vivo.

In one embodiment, the first plurality of selected library members is cloned or amplified on episomally replicable vectors, a multiplicity of said vectors is transferred into a cell and homologously recombined to form shuffled library members in vivo.

In one embodiment, the first plurality of selected library members is not fragmented, but is cloned or amplified on an episomally replicable vector as a direct repeat, which each repeat comprising a distinct species of selected library member sequence, said vector is transferred into a cell and homologously recombined by intra-vector recombination to form shuffled library members in vivo.

In an embodiment, combinations of in vitro and in vivo shuffling are provided to enhance combinatorial diversity.

The present invention provides a method for generating libraries of displayed antibodies suitable for affinity interaction screening. The method comprises (1) obtaining a first plurality of selected library members comprising a displayed antibody and an associated polynucleotide encoding said displayed antibody, and obtaining said associated polynucleotides

or copies thereof, wherein said associated polynucleotides comprise a region of substantially identical variable region framework sequence, and (2) pooling and fragmenting said associated polynucleotides or copies to form fragments thereof

5   under conditions suitable for PCR amplification and thereby homologously recombining said fragments to form a shuffled pool of recombined polynucleotides comprising novel combinations of CDRs, whereby a substantial fraction (e.g., greater than 10 percent) of the recombined polynucleotides of said shuffled pool

10 comprise CDR combinations are not present in the first plurality of selected library members, said shuffled pool composing a library of displayed antibodies comprising CDR permutations and suitable for affinity interaction screening. Optionally, the shuffled pool is subjected to affinity screening to select

15 shuffled library members which bind to a predetermined epitope (antigen) and thereby selecting a plurality of selected shuffled library members. Optionally, the plurality of selected shuffled library members can be shuffled and screened iteratively, from 1 to about 1000 cycles or as desired until library members having

20 a desired binding affinity are obtained.

Accordingly, one aspect of the present invention provides a method for introducing one or more mutations into a template double-stranded polynucleotide, wherein the template double-stranded polynucleotide has been cleaved into random fragments of

25 a desired size, by adding to the resultant population of double-stranded fragments one or more single or double-stranded oligonucleotides, wherein said oligonucleotides comprise an area of identity and an area of heterology to the template polynucleotide; denaturing the resultant mixture of double-

30 stranded random fragments and oligonucleotides into single-stranded fragments; incubating the resultant population of single-stranded fragments with a polymerase under conditions which result in the annealing of said single-stranded fragments at regions of identity between the single-stranded fragments and

35 formation of a mutagenized double-stranded polynucleotide; and repeating the above steps as desired.

In another aspect the present invention is directed to a method of producing recombinant proteins having biological

activity by treating a sample comprising double-stranded template polynucleotides encoding a wild-type protein under conditions which provide for the cleavage of said template polynucleotides into random double-stranded fragments having a desired size; adding to the resultant population of random fragments one or more single or double-stranded oligonucleotides, wherein said oligonucleotides comprise areas of identity and areas of heterology to the template polynucleotide; denaturing the resultant mixture of double-stranded fragments and oligonucleotides into single-stranded fragments; incubating the resultant population of single-stranded fragments with a polymerase under conditions which result in the annealing of said single-stranded fragments at the areas of identity and formation of a mutagenized double-stranded polynucleotide; repeating the above steps as desired; and then expressing the recombinant protein from the mutagenized double-stranded polynucleotide.

A third aspect of the present invention is directed to a method for obtaining a chimeric polynucleotide by treating a sample comprising different double-stranded template polynucleotides wherein said different template polynucleotides contain areas of identity and areas of heterology under conditions which provide for the cleavage of said template polynucleotides into random double-stranded fragments of a desired size; denaturing the resultant random double-stranded fragments contained in the treated sample into single-stranded fragments; incubating the resultant single-stranded fragments with polymerase under conditions which provide for the annealing of the single-stranded fragments at the areas of identity and the formation of a chimeric double-stranded polynucleotide sequence comprising template polynucleotide sequences; and repeating the above steps as desired.

A fourth aspect of the present invention is directed to a method of replicating a template polynucleotide by combining in vitro single-stranded template polynucleotides with small random single-stranded fragments resulting from the cleavage and denaturation of the template polynucleotide, and incubating said mixture of nucleic acid fragments in the presence of a nucleic acid polymerase under conditions wherein a population of double-

stranded template polynucleotides is formed.

The invention also provides the use of polynucleotide
shuffling, *in vitro* and/or *in vivo* to shuffle polynculeotides
encoding polypeptides and/or polynucleotides comprising
5   transcriptional regulatory sequences.

The invention also provides the use of polynucleotide
shuffling to shuffle a population of viral genes (e.g., capsid
proteins, spike glycoproteins, polymerases, proteases, etc.) or
viral genomes (e.g., paramyxoviridae, orthomyxoviridae,
10   herpesviruses, retroviruses, reoviruses, rhinoviruses, etc.). In
an embodiment, the invnetion provides a method for shuffling
sequences encoding all or portions of immunogenic viral proteins
to generate novel combinations of epitopes as well as novel
epitopes created by recombination; such shuffled viral proteins
15   - may comprise epitopes or combinations of epitopes which are
likely to arise in the natural environment as a consequence of
viral evolution (e.g., such as recombination of influenza virus
strains).

The invention also provides a method suitable for
20   shuffling polynucleotide sequences for generating gene therapy
vectors and replication-defective gene therapy constructs, such
as may be used for human gene therapy, including but not limited
to vaccination vectors for DNA-based vaccination, as well as
anti-neoplastic gene therapy and other gene therapy formats.

25

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic diagram comparing mutagenic
shuffling over error-prone PCR; (a) the initial library; (b) pool
of selected sequences in first round of affinity selection; (d)
30   in vitro recombination of the selected sequences ('shuffling');
(f) pool of selected sequences in second round of affinity
selection after shuffling; (c) error-prone PCR; (e) pool of
selected sequences in second round of affinity selection after
error-prone PCR.

35   Figure 2 illustrates the reassembly of a 1.0 kb LacZ alpha
gene fragment from 10-50 bp random fragments. (a) Photograph of
a gel of PCR amplified DNA fragment having the LacZ alpha gene.
(b) Photograph of a gel of DNA fragments after digestion with

DNAseI. (c) Photograph of a gel of DNA fragments of 10-50 bp purified from the digested LacZ alpha gene DNA fragment; (d) Photograph of a gel of the 10-50 bp DNA fragments after the indicated number of cycles of DNA reassembly; (e) Photograph of a gel of the recombination mixture after amplification by PCR with primers.

Figure 3 is a schematic illustration of the LacZ alpha gene stop codon mutants and their DNA sequences. The boxed regions are heterologous areas, serving as markers. The stop codons are located in smaller boxes or underlined. "+" indicates a wild-type gene and "-" indicates a mutated area in the gene.

Figure 4 is a schematic illustration of the introduction or spiking of a synthetic oligonucleotide into the reassembly process of the LacZ alpha gene.

Figure 5 illustrates the regions of homology between a murine IL1-B gene (M) and a human IL1-B gene (H) with E. coli codon usage. Regions of heterology are boxed. The "_|⁻" indicate crossovers obtained upon the shuffling of the two genes.

Figure 6 is a schematic diagram of the antibody CDR shuffling model system using the scFv of anti-rabbit IgG antibody (A10B).

Figure 7 illustrates the observed frequency of occurrence of certain combinations of CDRs in the shuffled DNA of the scFv of anti-rabbit IgG antibody (A10B).

Figure 8 illustrates the improved avidity of the scFv anti-rabbit antibody after DNA shuffling and each cycle of selection.

Figure 9 schematically portrays pBR322-Sfi-BL-LA-Sfi and in vivo intraplasmidic recombination via direct repeats, as well as the rate of generation of ampicillin-resistant colonies by intraplasmidic recombination reconstituting a functional beta-lactamase gene.

Figure 10 schematically portrays pBR322-Sfi-2Bla-Sfi and in vivo intraplasmidic recombination via direct repeats, as well as the rate of generation of ampicillin-resistant colonies by intraplasmidic recombination reconstituting a functional beta-lactamase gene.

Figure 11 illustrates the method for testing the efficiency of multiple rounds of homologous recombination after the

introduction of polynucleotide fragments into cells for the generation of recombinant proteins.

Figure 12 schematically portrays generation of a library of vectors by shuffling cassettes at the following loci: promoter, leader peptide, terminator, selectable drug resistance gene, and origin of replication. The multiple parallel lines at each locus represents the multiplicity of cassettes for that cassette.

Figure 13 schematically shows some examples of cassettes suitable at various loci for constructing prokaryotic vector libraries by shuffling.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention relates to a method for nucleic acid molecule reassembly after random fragmentation and its application to mutagenesis of DNA sequences. Also described is a method for the production of nucleic acid fragments encoding mutant proteins having enhanced biological activity. In particular, the present invention also relates to a method of repeated cycles of mutagenesis, nucleic acid shuffling and selection which allow for the creation of mutant proteins having enhanced biological activity.

The present invention is directed to a method for generating a very large library of DNA, RNA or protein mutants. This method has particular advantages in the generation of related DNA fragments from which the desired nucleic acid fragment(s) may be selected. In particular the present invention also relates to a method of repeated cycles of mutagenesis, homologous recombination and selection which allow for the creation of mutant proteins having enhanced biological activity.

However, prior to discussing this invention in further detail, the following terms will first be defined.

Definitions

As used herein, the following terms have the following meanings:

The term "DNA reassembly" is used when recombination occurs between identical sequences.

By contrast, the term "DNA shuffling" is used herein to

indicate recombination between substantially homologous but non-identical sequences, in some embodiments DNA shuffling may involve crossover via nonhomologous recombination, such as via cre/lox and/or flp/frt systems and the like.

The term "amplification" means that the number of copies of a nucleic acid fragment is increased.

The term "identical" or "identity" means that two nucleic acid sequences have the same sequence or a complementary sequence. Thus, "areas of identity" means that regions or areas of a nucleic acid fragment or polynucleotide are identical or complementary to another polynucleotide or nucleic acid fragment.

The term "corresponds to" is used herein to mean that a polynucleotide sequence is homologous (i.e., is identical, not strictly evolutionarily related) to all or a portion of a reference polynucleotide sequence, or that a polypeptide sequence is identical to a reference polypeptide sequence. In contradistinction, the term "complementary to" is used herein to mean that the complementary sequence is homologous to all or a portion of a reference polynucleotide sequence. For illustration, the nucleotide sequence "TATAC" corresponds to a reference sequence "TATAC" and is complementary to a reference sequence "GTATA".

The following terms are used to describe the sequence relationships between two or more polynucleotides: "reference sequence", "comparison window", "sequence identity", "percentage of sequence identity", and "substantial identity". A "reference sequence" is a defined sequence used as a basis for a sequence comparison; a reference sequence may be a subset of a larger sequence, for example, as a segment of a full-length cDNA or gene sequence given in a sequence listing, such as a polynucleotide sequence of Fig. 1 or Fig. 2(b), or may comprise a complete cDNA or gene sequence. Generally, a reference sequence is at least 20 nucleotides in length, frequently at least 25 nucleotides in length, and often at least 50 nucleotides in length. Since two polynucleotides may each (1) comprise a sequence (i.e., a portion of the complete polynucleotide sequence) that is similar between the two polynucleotides, and (2) may further comprise a sequence that is divergent between the two polynucleotides, sequence

comparisons between two (or more) polynucleotides are typically
performed by comparing sequences of the two polynucleotides over
a "comparison window" to identify and compare local regions of
sequence similarity.

5   A "comparison window", as used herein, refers to a
conceptual segment of at least 20 contiguous nucleotide positions
wherein a polynucleotide sequence may be compared to a reference
sequence of at least 20 contiguous nucleotides and wherein the
portion of the polynucleotide sequence in the comparison window

10   may comprise additions or deletions (i.e., gaps) of 20 percent or
less as compared to the reference sequence (which does not
comprise additions or deletions) for optimal alignment of the two
sequences.   Optimal alignment of sequences for aligning a
comparison window may be conducted by the local homology

15   algorithm of Smith and Waterman (1981) Adv. Appl. Math. 2: 482,
by the homology alignment algorithm of Needleman and Wunsch
(1970) J. Mol. Biol. 48: 443, by the search for similarity method
of Pearson and Lipman (1988) Proc. Natl. Acad. Sci. (U.S.A.) 85:
2444, by computerized implementations of these algorithms (GAP,

20   BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software
Package Release 7.0, Genetics Computer Group, 575 Science Dr.,
Madison, WI), or by inspection, and the best alignment (i.e.,
resulting in the highest percentage of homology over the
comparison window) generated by the various methods is selected.

25   The term "sequence identity" means that two
polynucleotide sequences are identical (i.e., on a nucleotide-by-
nucleotide basis) over the window of comparison.   The term
"percentage of sequence identity" is calculated by comparing two
optimally aligned sequences over the window of comparison,

30   determining the number of positions at which the identical
nucleic acid base (e.g., A, T, C, G, U, or I) occurs in both
sequences to yield the number of matched positions, dividing the
number of matched positions by the total number of positions in
the window of comparison (i.e., the window size), and multiplying

35   the result by 100 to yield the percentage of sequence identity.
The terms "substantial identity" as used herein denotes a
characteristic of a polynucleotide sequence, wherein the
polynucleotide comprises a sequence that has at least 80 percent

sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison window of at least 20 nucleotide positions, frequently over a window of at least 25-50 nucleotides, wherein the percentage of sequence identity is calculated by comparing the reference sequence to the polynucleotide sequence which may include deletions or additions which total 20 percent or less of the reference sequence over the window of comparison.

Conservative amino acid substitutions refer to the interchangeability of residues having similar side chains. For example, a group of amino acids having aliphatic side chains is glycine, alanine, valine, leucine, and isoleucine; a group of amino acids having aliphatic-hydroxyl side chains is serine and threonine; a group of amino acids having amide-containing side chains is asparagine and glutamine; a group of amino acids having aromatic side chains is phenylalanine, tyrosine, and tryptophan; a group of amino acids having basic side chains is lysine, arginine, and histidine; and a group of amino acids having sulfur-containing side chains is cysteine and methionine. Preferred conservative amino acids substitution groups are: valine-leucine-isoleucine, phenylalanine-tyrosine, lysine-arginine, alanine-valine, and asparagine-glutamine.

The term "homologous" or "homeologous" means that one single-stranded nucleic acid sequence may hybridize to a complementary single-stranded nucleic acid sequence. The degree of hybridization may depend on a number of factors including the amount of identity between the sequences and the hybridization conditions such as temperature and salt concentration as discussed later. Preferably the region of identity is greater than about 5 bp, more preferably the region of identity is greater than 10 bp.

The term "heterologous" means that one single-stranded nucleic acid sequence is unable to hybridize to another single-stranded nucleic acid sequence or its complement. Thus areas of heterology means that nucleic acid fragments or polynucleotides have areas or regions in the sequence which are unable to hybridize to another nucleic acid or polynucleotide. Such

regions or areas are, for example, areas of mutations.

The term "cognate" as used herein refers to a gene sequence that is evolutionarily and functionally related between species. For example but not limitation, in the human genome, the human CD4 gene is the cognate gene to the mouse CD4 gene, since the sequences and structures of these two genes indicate that they are highly homologous and both genes encode a protein which functions in signaling T cell activation through MHC class II-restricted antigen recognition.

The term "wild-type" means that the nucleic acid fragment does not comprise any mutations. A "wild-type" protein means that the protein will be active at a level of activity found in nature and will comprise the amino acid sequence found in nature.

The term "related polynucleotides" means that regions or areas of the polynucleotides are identical and regions or areas of the polynucleotides are heterologous.

The term "chimeric polynucleotide" means that the polynucleotide comprises regions which are wild-type and regions which are mutated. It may also mean that the polynucleotide comprises wild-type regions from one polynucleotide and wild-type regions from another related polynucleotide.

The term "cleaving" means digesting the polynucleotide with enzymes or breaking the polynucleotide.

The term "population" as used herein means a collection of components such as polynucleotides, nucleic acid fragments or proteins. A "mixed population" means a collection of components which belong to the same family of nucleic acids or proteins (i.e. are related) but which differ in their sequence (i.e. are not identical) and hence in their biological activity.

The term "specific nucleic acid fragment" means a nucleic acid fragment having certain end points and having a certain nucleic acid sequence. Two nucleic acid fragments wherein one nucleic acid fragment has the identical sequence as a portion of the second nucleic acid fragment but different ends comprise two different specific nucleic acid fragments.

The term "mutations" means changes in the sequence of a wild-type nucleic acid sequence or changes in the sequence of a peptide. Such mutations may be point mutations such as

transitions or transversions. The mutations may be deletions, insertions or duplications.

In the polypeptide notation used herein, the lefthand direction is the amino terminal direction and the righthand direction is the carboxy-terminal direction, in accordance with standard usage and convention. Similarly, unless specified otherwise, the lefthand end of single-stranded polynucleotide sequences is the 5' end; the lefthand direction of double-stranded polynucleotide sequences is referred to as the 5' direction. The direction of 5' to 3' addition of nascent RNA transcripts is referred to as the transcription direction; sequence regions on the DNA strand having the same sequence as the RNA and which are 5' to the 5' end of the RNA transcript are referred to as "upstream sequences"; sequence regions on the DNA strand having the same sequence as the RNA and which are 3' to the 3' end of the coding RNA transcript are referred to as "downstream sequences".

The term "naturally-occurring" as used herein as applied to an object refers to the fact that an object can be found in nature. For example, a polypeptide or polynucleotide sequence that is present in an organism (including viruses) that can be isolated from a source in nature and which has not been intentionally modified by man in the laboratory is naturally-occurring. Generally, the term naturally-occurring refers to an object as present in a non-pathological (undiseased) individual, such as would be typical for the species.

The term "agent" is used herein to denote a chemical compound, a mixture of chemical compounds, an array of spatially localized compounds (e.g., a VLSIPS peptide array, polynucleotide array, and/or combinatorial small molecule array), a biological macromolecule, a bacteriophage peptide display library, a bacteriophage antibody (e.g., scFv) display library, a polysome peptide display library, or an extract made from biological materials such as bacteria, plants, fungi, or animal (particularly mammalian) cells or tissues. Agents are evaluated for potential activity as antineoplastics, anti-inflammatories, or apoptosis modulators by inclusion in screening assays described hereinbelow. Agents are evaluated for potential

activity as specific protein interaction inhibitors (i.e., an
agent which selectively inhibits a binding interaction between
two predetermined polypeptides but which does not substantially
interfere with cell viability) by inclusion in screening assays

5   described hereinbelow.

As used herein, "substantially pure" means an object
species is the predominant species present (i.e., on a molar
basis it is more abundant than any other individual
macromolecular species in the composition), and preferably a

10   substantially purified fraction is a composition wherein the
object species comprises at least about 50 percent (on a molar
basis) of all macromolecular species present. Generally, a
substantially pure composition will comprise more than about 80
to 90 percent of all macromolecular species present in the

15   composition. Most preferably, the object species is purified to
essential homogeneity (contaminant species cannot be detected in
the composition by conventional detection methods) wherein the
composition consists essentially of a single macromolecular
species. Solvent species, small molecules (<500 Daltons), and

20   elemental ion species are not considered macromolecular species.

As used herein the term "physiological conditions"
refers to temperature, pH, ionic strength, viscosity, and like
biochemical parameters which are compatible with a viable
organism, and/or which typically exist intracellularly in a

25   viable cultured yeast cell or mammalian cell. For example, the
intracellular conditions in a yeast cell grown under typical
laboratory culture conditions are physiological conditions.
Suitable in vitro reaction conditions for in vitro transcription
cocktails are generally physiological conditions. In general, in

30   vitro physiological conditions comprise 50-200 mM NaCl or KCl, pH
6.5-8.5, 20-45°C and 0.001-10 mM divalent cation (e.g., Mg^{++},
Ca^{++}); preferably about 150 mM NaCl or KCl, pH 7.2-7.6, 5 mM
divalent cation, and often include 0.01-1.0 percent nonspecific
protein (e.g., BSA). A non-ionic detergent (Tween, NP-40, Triton

35   X-100) can often be present, usually at about 0.001 to 2%,
typically 0.05-0.2% (v/v). Particular aqueous conditions may be
selected by the practitioner according to conventional methods.
For general guidance, the following buffered aqueous conditions

may be applicable: 10-250 mM NaCl, 5-50 mM Tris HCl, pH 5-8, with optional addition of divalent cation(s) and/or metal chelators and/or nonionic detergents and/or membrane fractions and/or antifoam agents and/or scintillants.

5   Specific hybridization is defined herein as the formation of hybrids between a first polynucleotide and a second polynucleotide (e.g., a polynucleotide having a distinct but substantially identical sequence to the first polynucleotide), wherein the first polynucleotide preferentially hybridizes to the
10   second polynucleotide under stringent hybridization conditions wherein substantially unrelated polynucleotide sequences do not form hybrids in the mixture.

As used herein, the term "single-chain antibody" refers to a polypeptide comprising a V_H domain and a V_L domain in
15   polypeptide linkage, generally linked via a spacer peptide (e.g., [Gly-Gly-Gly-Gly-Ser]$_x$), and which may comprise additional amino acid sequences at the amino- and/or carboxy- termini. For example, a single-chain antibody may comprise a tether segment for linking to the encoding polynucleotide. As an example, a
20   scFv is a single-chain antibody. Single-chain antibodies are generally proteins consisting of one or more polypeptide segments of at least 10 contiguous amino acids substantially encoded by genes of the immunoglobulin superfamily (e.g., see The Immunoglobulin Gene Superfamily, A.F. Williams and A.N. Barclay,
25   in Immunoglobulin Genes, T. Honjo, F.W. Alt, and T.H. Rabbitts, eds., (1989) Academic Press: San Diego, CA, pp.361-387, which is incorporated herein by reference), most frequently encoded by a rodent, non-human primate, avian, porcine, bovine, ovine, goat, or human heavy chain or light chain gene sequence. A functional
30   single-chain antibody generally contains a sufficient portion of an immunoglobulin superfamily gene product so as to retain the property of binding to a specific target molecule, typically a receptor or antigen (epitope).

As used herein, the term "complementarity-determining
35   region" and "CDR" refer to the art-recognized term as exemplified by the Kabat and Chothia CDR definitions also generally known as hypervariable regions or hypervariable loops (Chothia and Lesk (1987) J. Mol. Biol. 196: 901; Chothia et al. (1989) Nature 342:

877; E.A. Kabat et al., Sequences of Proteins of Immunological Interest (National Institutes of Health, Bethesda, MD) (1987); and Tramontano et al. (1990) J. Mol. Biol. 215: 175). Variable region domains typically comprise the amino-terminal

5   approximately 105-115 amino acids of a naturally-occurring immunoglobulin chain (e.g., amino acids 1-110), although variable domains somewhat shorter or longer are also suitable for forming single-chain antibodies.

An immunoglobulin light or heavy chain variable region

10   consists of a "framework" region interrupted by three hypervariable regions, also called CDR's. The extent of the framework region and CDR's have been precisely defined (see, "Sequences of Proteins of Immunological Interest," E. Kabat et al., 4th Ed., U.S. Department of Health and Human Services,

15   Bethesda, MD (1987)). The sequences of the framework regions of different light or heavy chains are relatively conserved within a species. As used herein, a "human framework region" is a framework region that is substantially identical (about 85% or more, usually 90-95% or more) to the framework region of a

20   naturally occurring human immunoglobulin. The framework region of an antibody, that is the combined framework regions of the constituent light and heavy chains, serves to position and align the CDR's. The CDR's are primarily responsible for binding to an epitope of an antigen.

25   As used herein, the term "variable segment" refers to a portion of a nascent peptide which comprises a random, pseudorandom, or defined kernal sequence. A variable segment can comprise both variant and invariant residue positions, and the degree of residue variation at a variant residue position may be

30   limited; both options are selected at the discretion of the practitioner. Typically, variable segments are about 5 to 20 amino acid residues in length (e.g., 8 to 10), although variable segments may be longer and may comprise antibody portions or receptor proteins, such as an antibody fragment, a nucleic acid

35   binding protein, a receptor protein, and the like.

As used herein, "random peptide sequence" refers to an amino acid sequence composed of two or more amino acid monomers and constructed by a stochastic or random process. A random

peptide can include framework or scaffolding motifs, which may comprise invariant sequences.

As used herein "random peptide library" refers to a set of polynucleotide sequences that encodes a set of random peptides, and to the set of random peptides encoded by those polynucleotide sequences, as well as the fusion proteins containing those random peptides.

As used herein, the term "pseudorandom" refers to a set of sequences that have limited variability, so that for example the degree of residue variability at one position is different than the degree of residue variability at another position, but any pseudorandom position is allowed some degree of residue variation, however circumscribed.

As used herein, the term "defined sequence framework" refers to a set of defined sequences that are selected on a nonrandom basis, generally on the basis of experimental data or structural data; for example, a defined sequence framework may comprise a set of amino acid sequences that are predicted to form a β-sheet structure or may comprise a leucine zipper heptad repeat motif, a zinc-finger domain, among other variations. A "defined sequence kernal" is a set of sequences which encompass a limited scope of variability. Whereas (1) a completely random 10-mer sequence of the 20 conventional amino acids can be any of $(20)^{10}$ sequences, and (2) a pseudorandom 10-mer sequence of the 20 conventional amino acids can be any of $(20)^{10}$ sequences but will exhibit a bias for certain residues at certain positions and/or overall, (3) a defined sequence kernal is a subset of sequences which is less that the maximum number of potential sequences if each residue position was allowed to be any of the allowable 20 conventional amino acids (and/or allowable unconventional amino/imino acids). A defined sequence kernal generally comprises variant and invariant residue positions and/or comprises variant residue positions which can comprise a residue selected from a defined subset of amino acid residues), and the like, either segmentally or over the entire length of the individual selected library member sequence. Defined sequence kernals can refer to either amino acid sequences or polynucleotide sequences. For illustration and not limitation,

the sequences $(NNK)_{10}$ and $(NNM)_{10}$, where N represents A, T, G, or C; K represents G or T; and M represents A or C, are defined sequence kernals.

As used herein "epitope" refers to that portion of an antigen or other macromolecule capable of forming a binding interaction that interacts with the variable region binding pocket of an antibody. Typically, such binding interaction is manifested as an intermolecular contact with one or more amino acid residues of a CDR.

As used herein, "receptor" refers to a molecule that has an affinity for a given ligand. Receptors can be naturally occurring or synthetic molecules. Receptors can be employed in an unaltered state or as aggregates with other species. Receptors can be attached, covalently or noncovalently, to a binding member, either directly or via a specific binding substance. Examples of receptors include, but are not limited to, antibodies, including monoclonal antibodies and antisera reactive with specific antigenic determinants (such as on viruses, cells, or other materials), cell membrane receptors, complex carbohydrates and glycoproteins, enzymes, and hormone receptors.

As used herein "ligand" refers to a molecule, such as a random peptide or variable segment sequence, that is recognized by a particular receptor. As one of skill in the art will recognize, a molecule (or macromolecular complex) can be both a receptor and a ligand. In general, the binding partner having a smaller molecular weight is referred to as the ligand and the binding partner having a greater molecular weight is referred to as a receptor.

As used herein, "linker" or "spacer" refers to a molecule or group of molecules that connects two molecules, such as a DNA binding protein and a random peptide, and serves to place the two molecules in a preferred configuration, e.g., so that the random peptide can bind to a receptor with minimal steric hindrance from the DNA binding protein.

As used herein, the term "operably linked" refers to a linkage of polynucleotide elements in a functional relationship. A nucleic acid is "operably linked" when it is placed into a

functional relationship with another nucleic acid sequence. For
instance, a promoter or enhancer is operably linked to a coding
sequence if it affects the transcription of the coding sequence.
Operably linked means that the DNA sequences being linked are
5   typically contiguous and, where necessary to join two protein
coding regions, contiguous and in reading frame.

Methodology
          Nucleic acid shuffling is a method for *in vitro* or *in vivo*
10   homologous recombination of pools of nucleic acid fragments or
polynucleotides. Mixtures of related nucleic acid sequences or
polynucleotides are randomly fragmented, and reassembled to yield
a library or mixed population of recombinant nucleic acid
molecules or polynucleotides.
15   In contrast to cassette mutagenesis, only shuffling and
error-prone PCR allow one to mutate a pool of sequences blindly
(without sequence information other than primers).
          The advantage of the mutagenic shuffling of this invention
over error-prone PCR alone for repeated selection can best be
20   explained with an example from antibody engineering. In Figure
1 is shown a schematic diagram of DNA shuffling as described
herein. The initial library can consist of related sequences of
diverse origin (i.e. antibodies from naive mRNA) or can be
derived by any type of mutagenesis (including shuffling) of a
25   single antibody gene. A collection of selected complementarity
determining regions ("CDRs") is obtained after the first round of
affinity selection (Fig.1). In the diagram the thick CDRs confer
onto the antibody molecule increased affinity for the antigen.
Shuffling allows the free combinatorial association of all of the
30   CDR1s with all of the CDR2s with all of the CDR3s, etc. (Fig.1).
          This method differs from PCR, in that it is an inverse chain
reaction. In PCR, the number of polymerase start sites and the
number of molecules grows exponentially. However, the sequence
of the polymerase start sites and the sequence of the molecules
35   remains essentially the same. In contrast, in nucleic acid
reassembly or shuffling of random fragments the number of start
sites and the number (but not size) of the random fragments
decreases over time. For fragments derived from whole plasmids

the theoretical endpoint is a single, large concatemeric molecule.

Since cross-overs occur at regions of homology, recombination will primarily occur between members of the same sequence family. This discourages combinations of CDRs that are grossly incompatible (eg. directed against different epitopes of the same antigen). It is contemplated that multiple families of sequences can be shuffled in the same reaction. Further, shuffling conserves the relative order, such that, for example, CDR1 will not be found in the position of CDR2.

Rare shufflants will contain a large number of the best (eg. highest affinity) CDRs and these rare shufflants may be selected based on their superior affinity (Fig. 1).

CDRs from a pool of 100 different selected antibody sequences can be permutated in up to 100^6 different ways. This large number of permutations cannot be represented in a single library of DNA sequences. Accordingly, it is contemplated that multiple cycles of DNA shuffling and selection may be required depending on the length of the sequence and the sequence diversity desired.

Error-prone PCR, in contrast, keeps all the selected CDRs in the same relative sequence (Fig. 1), generating a much smaller mutant cloud.

The template polynucleotide which may be used in the methods of this invention may be DNA or RNA. It may be of various lengths depending on the size of the gene or DNA fragment to be recombined or reassembled. Preferably the template polynucleotide is from 50 bp to 50 kb. It is contemplated that entire vectors containing the nucleic acid encoding the protein of interest can be used in the methods of this invention, and in fact have been successfully used.

The template polynucleotide may be obtained by amplification using the PCR reaction (U.S. Patent No. 4,683,202 and 4,683,195) or other amplification or cloning methods. However, the removal of free primers from the PCR product before fragmentation provides a more efficient result. Failure to adequately remove the primers can lead to a low frequency of crossover clones.

The template polynucleotide often should be double-stranded.

A double-stranded nucleic acid molecule is required to ensure that regions of the resulting single-stranded nucleic acid fragments are complementary to each other and thus can hybridize to form a double-stranded molecule.

5   It is contemplated that single-stranded or double-stranded nucleic acid fragments having regions of identity to the template polynucleotide and regions of heterology to the template polynucleotide may be added to the template polynucleotide at this step. It is also contemplated that two different but related

10   polynucleotide templates can be mixed at this step.

The double-stranded polynucleotide template and any added double-or single-stranded fragments are randomly digested into fragments of from about 5 bp to 5 kb or more. Preferably the size of the random fragments is from about 10 bp to 1000 bp, more

15   preferably the size of the DNA fragments is from about 20 bp to 500 bp.

Alternatively, it is also contemplated that double-stranded nucleic acid having multiple nicks may be used in the methods of this invention. A nick is a break in one strand of the double-

20   stranded nucleic acid. The distance between such nicks is preferably 5 bp to 5 kb, more preferably between 10 bp to 1000 bp.

The nucleic acid fragment may be digested by a number of different methods. The nucleic acid fragment may be digested

25   with a nuclease, such as DNAseI or RNAse. The nucleic acid may be randomly sheared by the method of sonication or by passage through a tube having a small orifice.

It is also contemplated that the nucleic acid may also be partially digested with one or more restriction enzymes, such

30   that certain points of cross-over may be retained statistically.

The concentration of any one specific nucleic acid fragment will not be greater than 1% by weight of the total nucleic acid, more preferably the concentration of any one specific nucleic acid sequence will not be greater than 0.1% by weight of the

35   total nucleic acid.

The number of different specific nucleic acid fragments in the mixture will be at least about 100, preferably at least about 500, and more preferably at least about 1000.

At this step single-stranded or double-stranded nucleic acid fragments, either synthetic or natural, may be added to the random double-stranded nucleic acid fragments in order to increase the heterogeneity of the mixture of nucleic acid fragments.

It is also contemplated that populations of double-stranded randomly broken nucleic acid fragments may be mixed or combined at this step.

Where insertion of mutations into the template polynucleotide is desired, single-stranded or double-stranded nucleic acid fragments having a region of identity to the template polynucleotide and a region of heterology to the template polynucleotide may be added in a 20 fold excess by weight as compared to the total nucleic acid, more preferably the single-stranded nucleic acid fragments may be added in a 10 fold excess by weight as compared to the total nucleic acid.

Where a mixture of different but related template polynucleotides is desired, populations of nucleic acid fragments from each of the templates may be combined at a ratio of less than about 1:100, more preferably the ratio is less than about 1:40. For example, a backcross of the wild-type polynucleotide with a population of mutated polynucleotide may be desired to eliminate neutral mutations (e.g., mutations yielding an insubstantial alteration in the phenotypic property being selected for). In such an example, the ratio of randomly digested wild-type polynucleotide fragments which may be added to the randomly digested mutant polynucleotide fragments is approximately 1:1 to about 100:1, and more preferably from 1:1 to 40:1.

The mixed population of random nucleic acid fragments are denatured to form single-stranded nucleic acid fragments and then reannealed. Only those single-stranded nucleic acid fragments having regions of homology with other single-stranded nucleic acid fragments will reanneal.

The random nucleic acid fragments may be denatured by heating. One skilled in the art could determine the conditions necessary to completely denature the double stranded nucleic acid. Preferably the temperature is from 80 °C to 100 °C, more

preferably the temperature is from 90 °C to 96 °C. Other methods which may be used to denature the nucleic acid fragments include pressure (36) and pH.

The nucleic acid fragments may be reannealed by cooling. Preferably the temperature is from 20 °C to 75 °C, more preferably the temperature is from 40 °C to 65 °C. If a high frequency of crossovers is needed based on an average of only 4 consecutive bases of homology, recombination can be forced by using a low annealing temperature, although the process becomes more difficult. The degree of renaturation which occurs will depend on the degree of homology between the population of single-stranded nucleic acid fragments.

Renaturation can be accelerated by the addition of polyethylene glycol ("PEG") or salt. The salt concentration is preferably from 0 mM to 200 mM, more preferably the salt concentration is from 10 mM to 100 mM. The salt may be KCl or NaCl. The concentration of PEG is preferably from 0% to 20%, more preferably from 5% to 10%.

The annealed nucleic acid fragments are next incubated in the presence of a nucleic acid polymerase and dNTP's (i.e. dATP, dCTP, dGTP and dTTP). The nucleic acid polymerase may be the Klenow fragment, the Taq polymerase or any other DNA polymerase known in the art.

The approach to be used for the assembly depends on the minimum degree of homology that should still yield crossovers. If the areas of identity are large, Taq polymerase can be used with an annealing temperature of between 45-65°C. If the areas of identity are small, Klenow polymerase can be used with an annealing temperature of between 20-30°C. One skilled in the art could vary the temperature of annealing to increase the number of cross-overs achieved.

The polymerase may be added to the random nucleic acid fragments prior to annealing, simultaneously with annealing or after annealing.

The cycle of denaturation, renaturation and incubation in the presence of polymerase is referred to herein as shuffling or reassembly of the nucleic acid. This cycle is repeated for a desired number of times. Preferably the cycle is repeated from

2 to 50 times, more preferably the sequence is repeated from 10 to 40 times.

The resulting nucleic acid is a larger double-stranded polynucleotide of from about 50 bp to about 100 kb, preferably the larger polynucleotide is from 500 bp to 50 kb.

This larger polynucleotide fragment may contain a number of copies of a nucleic acid fragment having the same size as the template polynucleotide in tandem. This concatemeric fragment is then digested into single copies of the template polynucleotide. The result will be a population of nucleic acid fragments of approximately the same size as the template polynucleotide. The population will be a mixed population where single or double-stranded nucleic acid fragments having an area of identity and an area of heterology have been added to the template polynucleotide prior to shuffling.

These fragment are then cloned into the appropriate vector and the ligation mixture used to transform bacteria.

It is contemplated that the single nucleic acid fragments may be obtained from the larger concatemeric nucleic acid fragment by amplification of the single nucleic acid fragments prior to cloning by a variety of methods including PCR (U.S. Patent No. 4,683,195 and 4,683,202) rather than by digestion of the concatemer.

The vector used for cloning is not critical provided that it will accept a DNA fragment of the desired size. If expression of the DNA fragment is desired, the cloning vehicle should further comprise transcription and translation signals next to the site of insertion of the DNA fragment to allow expression of the DNA fragment in the host cell. Preferred vectors include the pUC series and the pBR series of plasmids.

The resulting bacterial population will include a number of recombinant DNA fragments having random mutations. This mixed population may be tested to identify the desired recombinant nucleic acid fragment. The method of selection will depend on the DNA fragment desired.

For example, if a DNA fragment which encodes for a protein with increased binding efficiency to a ligand is desired, the proteins expressed by each of the DNA fragments in the population

or library may be tested for their ability to bind to the ligand
by methods known in the art (i.e. panning, affinity
chromatography). If a DNA fragment which encodes for a protein
with increased drug resistance is desired, the proteins expressed
by each of the DNA fragments in the population or library may be
tested for their ability to confer drug resistance to the host
organism. One skilled in the art, given knowledge of the desired
protein, could readily test the population to identify DNA
fragments which confer the desired properties onto the protein.

It is contemplated that one skilled in the art could use a
phage display system in which fragments of the protein are
expressed as fusion proteins on the phage surface (Pharmacia,
Milwaukee WI). The recombinant DNA molecules are cloned into the
phage DNA at a site which results in the transcription of a
fusion protein a portion of which is encoded by the recombinant
DNA molecule. The phage containing the recombinant nucleic acid
molecule undergoes replication and transcription in the cell.
The leader sequence of the fusion protein directs the transport
of the fusion protein to the tip of the phage particle. Thus the
fusion protein which is partially encoded by the recombinant DNA
molecule is displayed on the phage particle for detection and
selection by the methods described above.

It is further contemplated that a number of cycles of
nucleic acid shuffling may be conducted with nucleic acid
fragments from a subpopulation of the first population, which
subpopulation contains DNA encoding the desired recombinant
protein. In this manner, proteins with even higher binding
affinities or enzymatic activity could be achieved.

It is also contemplated that a number of cycles of nucleic
acid shuffling may be conducted with a mixture of wild-type
nucleic acid fragments and a subpopulation of nucleic acid from
the first or subsequent rounds of nucleic acid shuffling in order
to remove any silent mutations from the subpopulation.

Any source of nucleic acid, in purified form can be utilized
as the starting nucleic acid. Thus the process may employ DNA or
RNA including messenger RNA, which DNA or RNA may be single or
double stranded. In addition, a DNA-RNA hybrid which contains
one strand of each may be utilized. The nucleic acid sequence

may be of various lengths depending on the size of the nucleic
acid sequence to be mutated. Preferably the specific nucleic
acid sequence is from 50 to 50000 base pairs. It is contemplated
that entire vectors containing the nucleic acid encoding the
5   protein of interest may be used in the methods of this invention.

        The nucleic acid may be obtained from any source, for
example, from plasmids such a pBR322, from cloned DNA or RNA or
from natural DNA or RNA from any source including bacteria,
10   yeast, viruses and higher organisms such as plants or animals.
DNA or RNA may be extracted from blood or tissue material. The
template polynucleotide may be obtained by amplification using
the polynucleotide chain reaction (PCR) (U.S. Patent no.
4,683,202 and 4,683,195). Alternatively, the polynucleotide may
15   be present in a vector present in a cell and sufficient nucleic
acid may be obtained by culturing the cell and extracting the
nucleic acid from the cell by methods known in the art.
        Any specific nucleic acid sequence can be used to produce
the population of mutants by the present process. It is only
20   necessary that a small population of mutant sequences of the
specific nucleic acid sequence exist or be created prior to the
present process.
        The initial small population of the specific nucleic acid
sequences having mutations may be created by a number of
25   different methods. Mutations may be created by error-prone PCR.
Error-prone PCR uses low-fidelity polymerization conditions to
introduce a low level of point mutations randomly over a long
sequence. Alternatively, mutations can be introduced into the
template polynucleotide by oligonucleotide-directed mutagenesis.
30   In oligonucleotide-directed mutagenesis, a short sequence of the
polynucleotide is removed from the polynucleotide using
restriction enzyme digestion and is replaced with a synthetic
polynucleotide in which various bases have been altered from the
original sequence. The polynucleotide sequence can also be
35   altered by chemical mutagenesis. Chemical mutagens include, for
example, sodium bisulfite, nitrous acid, hydroxylamine, hydrazine
or formic acid. Other agents which are analogues of nucleotide
precursors include nitrosoguanidine, 5-bromouracil, 2-

aminopurine, or acridine. Generally, these agents are added to
the PCR reaction in place of the nucleotide precursor thereby
mutating the sequence. Intercalating agents such as proflavine,
acriflavine, quinacrine and the like can also be used. Random

5   mutagenesis of the polynucleotide sequence can also be achieved
by irradiation with X-rays or ultraviolet light. Generally,
plasmid DNA or DNA fragments so mutagenized are introduced into
E. coli and propagated as a pool or library of mutant plasmids.

Alternatively the small mixed population of specific nucleic

10   acids may be found in nature in that they may consist of
different alleles of the same gene or the same gene from
different related species (i.e., cognate genes). Alternatively,
they may be related DNA sequences found within one species, for
example, the immunoglobulin genes.

15   Once the mixed population of the specific nucleic acid
sequences is generated, the polynucleotides can be used directly
or inserted into an appropriate cloning vector, using techniques
well-known in the art.

The choice of vector depends on the size of the

20   polynucleotide sequence and the host cell to be employed in the
methods of this invention. The templates of this invention may
be plasmids, phages, cosmids, phagemids, viruses (e.g.,
retroviruses, parainfluenzavirus, herpesviruses, reoviruses,
paramyxoviruses, and the like), or selected portions thereof

25   (e.g., coat protein, spike glycoprotein, capsid protein). For
example, cosmids and phagemids are preferred where the specific
nucleic acid sequence to be mutated is larger because these
vectors are able to stably propagate large nucleic acid
fragments.

30   If the mixed population of the specific nucleic acid
sequence is cloned into a vector it can be clonally amplified by
inserting each vector into a host cell and allowing the host cell
to amplify the vector. This is referred to as clonal
amplification because while the absolute number of nucleic acid

35   sequences increases, the number of mutants does not increase.
Utility

The DNA shuffling method of this invention can be performed
blindly on a pool of unknown sequences. By adding to the

reassembly mixture oligonucleotides (with ends that are homologous to the sequences being reassembled) any sequence mixture can be incorporated at any specific position into another sequence mixture. Thus, it is contemplated that mixtures of synthetic oligonucleotides, PCR fragments or even whole genes can be mixed into another sequence library at defined positions. The insertion of one sequence (mixture) is independent from the insertion of a sequence in another part of the template. Thus, the degree of recombination, the homology required, and the diversity of the library can be independently and simultaneously varied along the length of the reassembled DNA.

This approach of mixing two genes may be useful for the humanization of antibodies from murine hybridomas. The approach of mixing two genes or inserting mutant sequences into genes may be useful for any therapeutically used protein, for example, interleukin I, antibodies, tPA, growth hormone, etc. The approach may also be useful in any nucleic acid for example, promoters or introns or 3' untranslated region or 5' untranslated regions of genes to increase expression or alter specificity of expression of proteins. The approach may also be used to mutate ribozymes or aptamers.

Shuffling requires the presence of homologous regions separating regions of diversity. Scaffold-like protein structures may be particularly suitable for shuffling. The conserved scaffold determines the overall folding by self-association, while displaying relatively unrestricted loops that mediate the specific binding. Examples of such scaffolds are the immunoglobulin beta-barrel, and the four-helix bundle (24). This shuffling can be used to create scaffold-like proteins with various combinations of mutated sequences for binding.

In Vitro Shuffling

The equivalents of some standard genetic matings may also be performed by shuffling *in vitro*. For example, a 'molecular backcross' can be performed by repeated mixing of the mutant's nucleic acid with the wild-type nucleic acid while selecting for the mutations of interest. As in traditional breeding, this approach can be used to combine phenotypes from different sources

into a background of choice. It is useful, for example, for the removal of neutral mutations that affect unselected characteristics (i.e. immunogenicity). Thus it can be useful to determine which mutations in a protein are involved in the enhanced biological activity and which are not, an advantage which cannot be achieved by error-prone mutagenesis or cassette mutagenesis methods.

Large, functional genes can be assembled correctly from a mixture of small random fragments. This reaction may be of use for the reassembly of genes from the highly fragmented DNA of fossils (25). In addition random nucleic acid fragments from fossils may be combined with nucleic acid fragments from similar genes from related species.

It is also contemplated that the method of this invention can be used for the *in vitro* amplification of a whole genome from a single cell as is needed for a variety of research and diagnostic applications. DNA amplification by PCR is in practice limited to a length of about 40 kb. Amplification of a whole genome such as that of *E. coli* (5,000 kb) by PCR would require about 250 primers yielding 125 forty kb fragments. This approach is not practical due to the unavailability of sufficient sequence data. On the other hand, random digestion of the genome with DNAseI, followed by gel purification of small fragments will provide a multitude of possible primers. Use of this mix of random small fragments as primers in a PCR reaction alone or with the whole genome as the template should result in an inverse chain reaction with the theoretical endpoint of a single concatemer containing many copies of the genome.

100 fold amplification in the copy number and an average fragment size of greater than 50 kb may be obtained when only random fragments are used (see Example 2). It is thought that the larger concatemer is generated by overlap of many smaller fragments. The quality of specific PCR products obtained using synthetic primers will be indistinguishable from the product obtained from unamplified DNA. It is expected that this approach will be useful for the mapping of genomes.

The polynucleotide to be shuffled can be fragmented randomly or non-randomly, at the discretion of the practitioner.

In Vivo Shuffling

In an embodiment of *in vivo* shuffling, the mixed population of the specific nucleic acid sequence is introduced into bacterial or eukaryotic cells under conditions such that at least two different nucleic acid sequences are present in each host cell. The fragments can be introduced into the host cells by a variety of different methods. The host cells can be transformed with the fragments using methods known in the art, for example treatment with calcium chloride. If the fragments are inserted into a phage genome, the host cell can be transfected with the recombinant phage genome having the specific nucleic acid sequences. Alternatively, the nucleic acid sequences can be introduced into the host cell using electroporation, transfection, lipofection, biolistics, conjugation, and the like.

In general, in this embodiment, the specific nucleic acids sequences will be present in vectors which are capable of stably replicating the sequence in the host cell. In addition, it is contemplated that the vectors will encode a marker gene such that host cells having the vector can be selected. This ensures that the mutated specific nucleic acid sequence can be recovered after introduction into the host cell. However, it is contemplated that the entire mixed population of the specific nucleic acid sequences need not be present on a vector sequence. Rather only a sufficient number of sequences need be cloned into vectors to ensure that after introduction of the fragments into the host cells each host cell contains one vector having at least one specific nucleic acid sequence present therein. It is also contemplated that rather than having a subset of the population of the specific nucleic acids sequences cloned into vectors, this subset may be already stably integrated into the host cell.

It has been found that when two fragments which have regions of identity are inserted into the host cells homologous recombination occurs between the two fragments. Such recombination between the two mutated specific nucleic acid sequences will result in the production of double or triple mutants in some situations.

It has also been found that the frequency of recombination is increased if some of the mutated specific nucleic acid

sequences are present on linear nucleic acid molecules.
Therefore, in a preferred embodiment, some of the specific
nucleic acid sequences are present on linear nucleic acid
fragments.

5   After transformation, the host cell transformants are placed
under selection to identify those host cell transformants which
contain mutated specific nucleic acid sequences having the
qualities desired. For example, if increased resistance to a
particular drug is desired then the transformed host cells may be
10   subjected to increased concentrations of the particular drug and
those transformants producing mutated proteins able to confer
increased drug resistance will be selected. If the enhanced
ability of a particular protein to bind to a receptor is desired,
then expression of the protein can be induced from the
15   transformants and the resulting protein assayed in a ligand
binding assay by methods known in the art to identify that subset
of the mutated population which shows enhanced binding to the
ligand. Alternatively, the protein can be expressed in another
system to ensure proper processing.

20   Once a subset of the first recombined specific nucleic acid
sequences (daughter sequences) having the desired characteristics
are identified, they are then subject to a second round of
recombination.

           In the second cycle of recombination, the recombined
25   specific nucleic acid sequences may be mixed with the original
mutated specific nucleic acid sequences (parent sequences) and
the cycle repeated as described above. In this way a set of
second recombined specific nucleic acids sequences can be
identified which have enhanced characteristics or encode for
30   proteins having enhanced properties. This cycle can be repeated
a number of times as desired.

           It is also contemplated that in the second or subsequent
recombination cycle, a backcross can be performed. A molecular
backcross can be performed by mixing the desired specific nucleic
35   acid sequences with a large number of the wild-type sequence,
such that at least one wild-type nucleic acid sequence and a
mutated nucleic acid sequence are present in the same host cell
after transformation. Recombination with the wild-type specific

nucleic acid sequence will eliminate those neutral mutations that
may affect unselected characteristics such as immunogenicity but
not the selected characteristics.

In another embodiment of this invention, it is contemplated
that during the first round a subset of the specific nucleic acid
sequences can be fragmented prior to introduction into the host
cell. The size of the fragments must be large enough to contain
some regions of identity with the other sequences so as to
homologously recombine with the other sequences. The size of the
fragments will range from 0.03 kb to 100 kb more preferably from
0.2 kb to 10 kb. It is also contemplated that in subsequent
rounds, all of the specific nucleic acid sequences other than the
sequences selected from the previous round may be cleaved into
fragments prior to introduction into the host cells.

. Fragmentation of the sequences can be accomplished by a
variety of method known in the art. The sequences can be
randomly fragmented or fragmented at specific sites in the
nucleic acid sequence. Random fragments can be obtained by
breaking the nucleic acid or exposing it to harsh physical
treatment (e.g., shearing or irradiation) or harsh chemical
agents (e.g., by free radicals; metal ions; acid treatment to
depurinate and cleave). Random fragments can also be obtained,
in the case of DNA by the use of DNase or like nuclease. The
sequences can be cleaved at specific sites by the use of
restriction enzymes. The fragmented sequences can be single-
stranded or double-stranded. If the sequences were originally
single-stranded they can be denatured with heat, chemicals or
enzymes prior to insertion into the host cell. The reaction
conditions suitable for separating the strands of nucleic acid
are well known in the art.

The steps of this process can be repeated indefinitely,
being limited only by the number of possible mutants which can be
achieved. After a certain number of cycles, all possible mutants
will have been achieved and further cycles are redundant.

In an embodiment the same mutated template nucleic acid is
repeatedly recombined and the resulting recombinants selected for
the desired characteristic.

. Therefore, the initial pool or population of mutated

template nucleic acid is cloned into a vector capable of replicating in a bacteria such as E. coli. The particular vector is not essential, so long as it is capable of autonomous replication in E. coli. In a preferred embodiment, the vector is designed to allow the expression and production of any protein encoded by the mutated specific nucleic acid linked to the vector. It is also preferred that the vector contain a gene encoding for a selectable marker.

The population of vectors containing the pool of mutated nucleic acid sequences is introduced into the E. coli host cells. The vector nucleic acid sequences may be introduced by transformation, transfection or infection in the case of phage. The concentration of vectors used to transform the bacteria is such that a number of vectors is introduced into each cell. Once present in the cell, the efficiency of homologous recombination is such that homologous recombination occurs between the various vectors. This results in the generation of mutants (daughters) having a combination of mutations which differ from the original parent mutated sequences.

The host cells are then clonally replicated and selected for the marker gene present on the vector. Only those cells having a plasmid will grow under the selection.

The host cells which contain a vector are then tested for the presence of favorable mutations. Such testing may consist of placing the cells under selective pressure, for example, if the gene to be selected is an improved drug resistance gene. If the vector allows expression of the protein encoded by the mutated nucleic acid sequence, then such selection may include allowing expression of the protein so encoded, isolation of the protein and testing of the protein to determine whether, for example, it binds with increased efficiency to the ligand of interest.

Once a particular daughter mutated nucleic acid sequence has been identified which confers the desired characteristics, the nucleic acid is isolated either already linked to the vector or separated from the vector. This nucleic acid is then mixed with the first or parent population of nucleic acids and the cycle is repeated.

It has been shown that by this method nucleic acid sequences

having enhanced desired properties can be selected.

In an alternate embodiment, the first generation of mutants are retained in the cells and the parental mutated sequences are added again to the cells. Accordingly, the first cycle of Embodiment I is conducted as described above. However, after the daughter nucleic acid sequences are identified, the host cells containing these sequences are retained.

The parent mutated specific nucleic acid population, either as fragments or cloned into the same vector is introduced into the host cells already containing the daughter nucleic acids. Recombination is allowed to occur in the cells and the next generation of recombinants, or granddaughters are selected by the methods described above.

This cycle can be repeated a number of times until the nucleic acid or peptide having the desired characteristics is obtained. It is contemplated that in subsequent cycles, the population of mutated sequences which are added to the preferred mutants may come from the parental mutants or any subsequent generation.

In an alternative embodiment, the invention provides a method of conducting a "molecular" backcross of the obtained recombinant specific nucleic acid in order to eliminate any neutral mutations. Neutral mutations are those mutations which do not confer onto the nucleic acid or peptide the desired properties. Such mutations may however confer on the nucleic acid or peptide undesirable characteristics. Accordingly, it is desirable to eliminate such neutral mutations. The method of this invention provide a means of doing so.

In this embodiment, after the mutant nucleic acid, having the desired characteristics, is obtained by the methods of the embodiments, the nucleic acid, the vector having the nucleic acid or the host cell containing the vector and nucleic acid is isolated.

The nucleic acid or vector is then introduced into the host cell with a large excess of the wild-type nucleic acid. The nucleic acid of the mutant and the nucleic acid of the wild-type sequence are allowed to recombine. The resulting recombinants are placed under the same selection as the mutant nucleic acid.

Only those recombinants which retained the desired
characteristics will be selected. Any silent mutations which do
not provide the desired characteristics will be lost through
recombination with the wild-type DNA. This cycle can be repeated

5   a number of times until all of the silent mutations are
eliminated.

Thus the methods of this invention can be used in a
molecular backcross to eliminate unnecessary or silent mutations.

10   Utility

The in vivo recombination method of this invention can be
performed blindly on a pool of unknown mutants or alleles of a
specific nucleic acid fragment or sequence. However, it is not
necessary to know the actual DNA or RNA sequence of the specific

15   nucleic acid fragment.

The approach of using recombination within a mixed
population of genes can be useful for the generation of any
useful proteins, for example, interleukin I, antibodies, tPA,
growth hormone, etc. This approach may be used to generate

20   proteins having altered specificity or activity. The approach
may also be useful for the generation of mutant nucleic acid
sequences, for example, promoter regions, introns, exons,
enhancer sequences, 3' untranslated regions or 5' untranslated
regions of genes. Thus this approach may be used to generate

25   genes having increased rates of expression. This approach may
also be useful in the study of repetitive DNA sequences.
Finally, this approach may be useful to mutate ribozymes or
aptamers.

Scaffold-like regions separating regions of diversity in

30   proteins may be particularly suitable for the methods of this
invention.1 The conserved scaffold determines the overall
folding by self-association, while displaying relatively
unrestricted loops that mediate the specific binding. Examples
of such scaffolds are the immunoglobulin beta barrel, and the

35   four-helix bundle. The methods of this invention can be used to
create scaffold-like proteins with various combinations of
mutated sequences for binding.

The equivalents of some standard genetic matings may also be

performed by the methods of this invention. For example, a
"molecular" backcross can be performed by repeated mixing of the
mutant's nucleic acid with the wild-type nucleic acid while
selecting for the mutations of interest. As in traditional
5   breeding, this approach can be used to combine phenotypes from
different sources into a background of choice. It is useful, for
example, for the removal of neutral mutations that affect
unselected characteristics (i.e. immunogenicity). Thus it can be
useful to determine which mutations in a protein are involved in
10   the enhanced biological activity and which are not.

Peptide Display Methods

The present method can be used to shuffle, by _in vitro_
and/or _in vivo_ recombination by any of the disclosed methods, and
15   in any combination, polynucleotide sequences selected by peptide
display methods, wherein an associated polynucleotide encodes a
displayed peptide which is screened for a phenotype (e.g., for
affinity for a predetermined receptor (ligand).

An increasingly important aspect of biopharmaceutical
20   drug development and molecular biology is the identification of
peptide structures, including the primary amino acid sequences,
of peptides or peptidomimetics that interact with biological
macromolecules. One method of identifying peptides that possess
a desired structure or functional property, such as binding to a
25   predetermined biological macromolecule (e.g., a receptor),
involves the screening of a large library or peptides for
individual library members which possess the desired structure or
functional property conferred by the amino acid sequence of the
peptide.

30   In addition to direct chemical synthesis methods for
generating peptide libraries, several recombinant DNA methods
also have been reported. One type involves the display of a
peptide sequence, antibody, or other protein on the surface of a
bacteriophage particle or cell. Generally, in these methods each
35   bacteriophage particle or cell serves as an individual library
member displaying a single species of displayed peptide in
addition to the natural bacteriophage or cell protein sequences.
Each bacteriophage or cell contains the nucleotide sequence

information encoding the particular displayed peptide sequence; thus, the displayed peptide sequence can be ascertained by nucleotide sequence determination of an isolated library member.

A well-known peptide display method involves the presentation of a peptide sequence on the surface of a filamentous bacteriophage, typically as a fusion with a bacteriophage coat protein. The bacteriophage library can be incubated with an immobilized, predetermined macromolecule or small molecule (e.g., a receptor) so that bacteriophage particles which present a peptide sequence that binds to the immobilized macromolecule can be differentially partitioned from those that do not present peptide sequences that bind to the predetermined macromolecule. The bacteriophage particles (i.e., library members) which are bound to the immobilized macromolecule are then recovered and replicated to amplify the selected bacteriophage subpopulation for a subsequent round of affinity enrichment and phage replication. After several rounds of affinity enrichment and phage replication, the bacteriophage library members that are thus selected are isolated and the nucleotide sequence encoding the displayed peptide sequence is determined, thereby identifying the sequence(s) of peptides that bind to the predetermined macromolecule (e.g., receptor). Such methods are further described in PCT patent publication Nos. 91/17271, 91/18980, and 91/10818 and 92/08278.

The latter PCT publication describes a recombinant DNA method for the display of peptide ligands that involves the production of a library of fusion proteins with each fusion protein composed of a first polypeptide portion, typically comprising a variable sequence, that is available for potential binding to a predetermined macromolecule, and a second polypeptide portion that binds to DNA, such as the DNA vector encoding the individual fusion protein. When transformed host cells are cultured under conditions that allow for expression of the fusion protein, the fusion protein binds to the DNA vector encoding it. Upon lysis of the host cell, the fusion protein/vector DNA complexes can be screened against a predetermined macromolecule in much the same way as bacteriophage particles are screened in the phage-based display system, with

the replication and sequencing of the DNA vectors in the selected
fusion protein/vector DNA complexes serving as the basis for
identification of the selected library peptide sequence(s).

Other systems for generating libraries of peptides and
like polymers have aspects of both the recombinant and _in vitro_
chemical synthesis methods. In these hybrid methods, cell-free
enzymatic machinery is employed to accomplish the _in vitro_
synthesis of the library members (i.e., peptides or
polynucleotides). In one type of method, RNA molecules with the
ability to bind a predetermined protein or a predetermined dye
molecule were selected by alternate rounds of selection and PCR
amplification (Tuerk and Gold (1990) _Science_ 249: 505; Ellington
and Szostak (1990) _Nature_ 346: 818). A similar technique was
used to identify DNA sequences which bind a predetermined human
transcription factor (Thiesen and Bach (1990) _Nucleic Acids Res._
18: 3203; Beaudry and Joyce (1992) _Science_ 257; 635; PCT patent
publication Nos. 92/05258 and 92/14843). In a similar fashion,
the technique of _in vitro_ translation has been used to synthesize
proteins of interest and has been proposed as a method for
generating large libraries of peptides. These methods which rely
upon _in vitro_ translation, generally comprising stabilized
polysome complexes, are described further in PCT patent
publication Nos. 88/08453, 90/05785, 90/07003, 91/02076,
91/05058, and 92/02536. Applicants have described methods in
which library members comprise a fusion protein having a first
polypeptide portion with DNA binding activity and a second
polypeptide portion having the library member unique peptide
sequence; such methods are suitable for use in cell-free _in vitro_
selection formats, among others.

The displayed peptide sequences can be of varying
lengths, typically from 3-5000 amino acids long or longer,
frequently from 5-100 amino acids long, and often from about 8-15
amino acids long. A library can comprise library members having
varying lengths of displayed peptide sequence, or may comprise
library members having a fixed length of displayed peptide
sequence. Portions or all of the displayed peptide sequence(s)
can be random, pseudorandom, defined set kernal, fixed, or the
like. The present display methods include methods for _in vitro_

and _in_ _vivo_ display of single-chain antibodies, such as nascent
scFv on polysomes or scFv displayed on phage, which enable large-
scale screening of scFv libraries having broad diversity of
variable region sequences and binding specificities.

5   The present invention also provides random,
pseudorandom, and defined sequence framework peptide libraries
and methods for generating and screening those libraries to
identify useful compounds (e.g., peptides, including single-chain
antibodies) that bind to receptor molecules or epitopes of
10   interest or gene products that modify peptides or RNA in a
desired fashion. The random, pseudorandom, and defined sequence
framework peptides are produced from libraries of peptide library
members that comprise displayed peptides or displayed single-
chain antibodies attached to a polynucleotide template from which
15   the displayed peptide was synthesized. The mode of attachment
may vary according to the specific embodiment of the invention
selected, and can include encapsidation in a phage particle or
incorporation in a cell.

        A method of affinity enrichment allows a very large
20   library of peptides and single-chain antibodies to be screened
and the polynucleotide sequence encoding the desired peptide(s)
or single-chain antibodies to be selected. The polynucleotide
can then be isolated and shuffled to recombine combinatorially
the amino acid sequence of the selected peptide(s) (or
25   predetermined portions thereof) or single-chain antibodies (or
just V_H, V_L, or CDR portions thereof). Using these methods, one
can identify a peptide or single-chain antibody as having a
desired binding affinity for a molecule and can exploit the
process of shuffling to converge rapidly to a desired high-
30   affinity peptide or scFv. The peptide or antibody can then be
synthesized in bulk by conventional means for any suitable use
(e.g., as a therapeutic or diagnostic agent).

        A significant advantage of the present invention is
that no prior information regarding an expected ligand structure
35   is required to isolate peptide ligands or antibodies of interest.
The peptide identified can have biological activity, which is
meant to include at least specific binding affinity for a
selected receptor molecule and, in some instances, will further

include the ability to block the binding of other compounds, to stimulate or inhibit metabolic pathways, to act as a signal or messenger, to stimulate or inhibit cellular activity, and the like.

5   The present invention also provides a method for shuffling a pool of polynucleotide sequences selected by affinity screening a library of polysomes displaying nascent peptides (including single-chain antibodies) for library members which bind to a predetermined receptor (e.g., a mammalian proteinaceous
10   receptor such as, for example, a peptidergic hormone receptor, a cell surface receptor, an intracellular protein which binds to other protein(s) to form intracellular protein complexes such as heterodimers and the like) or epitope (e.g., an immobilized protein, glycoprotein, oligosaccharide, and the like).

15   Polynucleotide sequences selected in a first selection round (typically by affinity selection for binding to a receptor (e.g., a ligand) by any of these methods are pooled and the pool(s) is/are shuffled by in vitro and/or in vivo recombination to produce a shuffled pool comprising a population of recombined
20   selected polynucleotide sequences. The recombined selected polynucleotide sequences are subjected to at least one subsequent selection round. The polynucleotide sequences selected in the subsequent selection round(s) can be used directly, sequenced, and/or subjected to one or more additional rounds of shuffling
25   and subsequent selection. Selected sequences can also be backcrossed with polynucleotide sequences encoding neutral sequences (i.e., having insubstantial functional effect on binding), such as for example by backcrossing with a wild-type or naturally-occurring sequence substantially identical to a
30   selected sequence to produce native-like functional peptides, which may be less immunogenic. Generally, during backcrossing subsequent selection is applied to retain the property of binding to the predetermined receptor (ligand).

          Prior to or concomitant with the shuffling of selected
35   sequences, the sequences can be mutagenized. In one embodiment, selected library members are cloned in a prokaryotic vector (e.g., plasmid, phagemid, or bacteriophage) wherein a collection of individual colonies (or plaques) representing discrete library

members are produced. Individual selected library members can
then be manipulated (e.g., by site-directed mutagenesis, cassette
mutagenesis, chemical mutagenesis, PCR mutagenesis, and the like)
to generate a collection of library members representing a kernal

5   of sequence diversity based on the sequence of the selected
library member. The sequence of an individual selected library
member or pool can be manipulated to incorporate random mutation,
pseudorandom mutation, defined kernal mutation (i.e., comprising
variant and invariant residue positions and/or comprising variant

10   residue positions which can comprise a residue selected from a
defined subset of amino acid residues), codon-based mutation, and
the like, either segmentally or over the entire length of the
individual selected library member sequence. The mutagenized
selected library members are then shuffled by in vitro and/or in

15   vivo recombinatorial shuffling as disclosed herein.
            The invention also provides peptide libraries
comprising a plurality of individual library members of the
invention, wherein (1) each individual library member of said
plurality comprises a sequence produced by shuffling of a pool of

20   selected sequences, and (2) each individual library member
comprises a variable peptide segment sequence or single-chain
antibody segment sequence which is distinct from the variable
peptide segment sequences or single-chain antibody sequences of
other individual library members in said plurality (although some

25   library members may be present in more than one copy per library
due to uneven amplification, stochastic probability, or the
like).
            The invention also provides a product-by-process,
wherein selected polynucleotide sequences having (or encoding a

30   peptide having) a predetermined binding specificity are formed by
the process of: (1) screening a displayed peptide or displayed
single-chain antibody library against a predetermined receptor
(e.g., ligand) or epitope (e.g., antigen macromolecule) and
identifying and/or enriching library members which bind to the

35   predetermined receptor or epitope to produce a pool of selected
library members, (2) shuffling by recombination the selected
library members (or amplified or cloned copies thereof) which
binds the predetermined epitope and has been thereby isolated

and/or enriched from the library to generate a shuffled library, and (3) screening the shuffled library against the predetermined receptor (e.g., ligand) or epitope (e.g., antigen macromolecule) and identifying and/or enriching shuffled library members which bind to the predetermined receptor or epitope to produce a pool of selected shuffled library members.

Antibody Display and Screening Methods

The present method can be used to shuffle, by _in vitro_ and/or _in vivo_ recombination by any of the disclosed methods, and in any combination, polynucleotide sequences selected by antibody display methods, wherein an associated polynucleotide encodes a displayed antibody which is screened for a phenotype (e.g., for affinity for binding a predetermined antigen (ligand).

Various molecular genetic approaches have been devised to capture the vast immunological repertoire represented by the extremely large number of distinct variable regions which can be present in immunoglobulin chains. The naturally-occurring germline immunoglobulin heavy chain locus is composed of separate tandem arrays of variable (V) segment genes located upstream of a tandem array of diversity (D) segment genes, which are themselves located upstream of a tandem array of joining (J) region genes, which are located upstream of the constant (C_H) region genes. During B lymphocyte development, V-D-J rearrangement occurs wherein a heavy chain variable region gene (V_H) is formed by rearrangement to form a fused D-J segment followed by rearrangement with a V segment to form a V-D-J joined product gene which, if productively rearranged, encodes a functional variable region (V_H) of a heavy chain. Similarly, light chain loci rearrange one of several V segments with one of several J segments to form a gene encoding the variable region (V_L) of a light chain.

The vast repertoire of variable regions possible in immunoglobulins derives in part from the numerous combinatorial possibilities of joining V and J segments (and, in the case of heavy chain loci, D segments) during rearrangement in B cell development. Additional sequence diversity in the heavy chain variable regions arises from non-uniform rearrangements of the D

segments during V-D-J joining and from N region addition. Further, antigen-selection of specific B cell clones selects for higher affinity variants having nongermline mutations in one or both of the heavy and light chain variable regions; a phenomenon

5   referred to as "affinity maturation" or "affinity sharpening". Typically, these "affinity sharpening" mutations cluster in specific areas of the variable region, most commonly in the complementarity-determining regions (CDRs).

In order to overcome many of the limitations in

10 producing and identifying high-affinity immunoglobulins through antigen-stimulated B cell development (i.e., immunization), various prokaryotic expression systems have been developed that can be manipulated to produce combinatorial antibody libraries which may be screened for high-affinity antibodies to specific

15 antigens. Recent advances in the expression of antibodies in Escherichia coli and bacteriophage systems (see, "Alternative Peptide Display Methods", infra) have raised the possibility that virtually any specificity can be obtained by either cloning antibody genes from characterized hybridomas or by de novo

20 selection using antibody gene libraries (e.g., from Ig cDNA).

Combinatorial libraries of antibodies have been generated in bacteriophage lambda expression systems which may be screened as bacteriophage plaques or as colonies of lysogens (Huse et al. (1989) Science 246: 1275; Caton and Koprowski (1990)

25 Proc. Natl. Acad. Sci. (U.S.A.) 87: 6450; Mullinax et al (1990) Proc. Natl. Acad. Sci. (U.S.A.) 87: 8095; Persson et al. (1991) Proc. Natl. Acad. Sci. (U.S.A.) 88: 2432). Various embodiments of bacteriophage antibody display libraries and lambda phage expression libraries have been described (Kang et al. (1991)

30 Proc. Natl. Acad. Sci. (U.S.A.) 88: 4363; Clackson et al. (1991) Nature 352: 624; McCafferty et al. (1990) Nature 348: 552; Burton et al. (1991) Proc. Natl. Acad. Sci. (U.S.A.) 88: 10134; Hoogenboom et al. (1991) Nucleic Acids Res. 19: 4133; Chang et al. (1991) J. Immunol. 147: 3610; Breitling et al. (1991) Gene

35 104: 147; Marks et al. (1991) J. Mol. Biol. 222: 581; Barbas et al. (1992) Proc. Natl. Acad. Sci. (U.S.A.) 89: 4457; Hawkins and Winter (1992) J. Immunol. 22: 867; Marks et al. (1992) Biotechnology 10: 779; Marks et al. (1992) J. Biol. Chem. 267:

16007; Lowman et al (1991) Biochemistry 30: 10832; Lerner et al.
(1992) Science 258: 1313, incorporated herein by reference).
Typically, a bacteriophage antibody display library is screened
with a receptor (e.g., polypeptide, carbohydrate, glycoprotein,
nucleic acid) that is immobilized (e.g., by covalent linkage to
a chromatography resin to enrich for reactive phage by affinity
chromatography) and/or labeled (e.g., to screen plaque or colony
lifts).

        One particularly advantageous approach has been the use
of so-called single-chain fragment variable (scFv) libraries
(Marks et al. (1992) Biotechnology 10: 779; Winter G and Milstein
C (1991) Nature 349: 293; Clackson et al. (1991) op.cit.; Marks
et al. (1991) J. Mol. Biol. 222: 581; Chaudhary et al. (1990)
Proc. Natl. Acad. Sci. (USA) 87: 1066; Chiswell et al. (1992)
TIBTECH 10: 80; McCafferty et al. (1990) op.cit.; and Huston et
al. (1988) Proc. Natl. Acad. Sci. (USA) 85: 5879). Various
embodiments of scFv libraries displayed on bacteriophage coat
proteins have been described.

        Beginning in 1988, single-chain analogues of Fv
fragments and their fusion proteins have been reliably generated
by antibody engineering methods. The first step generally
involves obtaining the genes encoding V_H and V_L domains with
desired binding properties; these V genes may be isolated from a
specific hybridoma cell line, selected from a combinatorial
V-gene library, or made by V gene synthesis. The single-chain Fv
is formed by connecting the component V genes with an
oligonucleotide that encodes an appropriately designed linker
peptide, such as (Gly-Gly-Gly-Gly-Ser)$_3$ or equivalent linker
peptide(s). The linker bridges the C-terminus of the first V
region and N-terminus of the second, ordered as either
V_H-linker-V_L or V_L-linker-V_H. In principle, the scFv binding site
can faithfully replicate both the affinity and specificity of its
parent antibody combining site.

        Thus, scFv fragments are comprised of V_H and V_L domains
linked into a single polypeptide chain by a flexible linker
peptide. After the scFv genes are assembled, they are cloned
into a phagemid and expressed at the tip of the M13 phage (or
similar filamentous bacteriophage) as fusion proteins with the

bacteriophage pIII (gene 3) coat protein. Enriching for phage
expressing an antibody of interest is accomplished by panning the
recombinant phage displaying a population scFv for binding to a
predetermined epitope (e.g., target antigen, receptor).

5   The linked polynucleotide of a library member provides
the basis for replication of the library member after a screening
or selection procedure, and also provides the basis for the
determination, by nucleotide sequencing, of the identity of the
displayed peptide sequence or V_H and V_L amino acid sequence. The

10   displayed peptide(s) or single-chain antibody (e.g., scFv) and/or
its V_H and V_L domains or their CDRs can be cloned and expressed
in a suitable expression system. Often polynucleotides encoding
the isolated V_H and V_L domains will be ligated to polynucleotides
encoding constant regions (C_H and C_L) to form polynucleotides

15   encoding complete antibodies (e.g., chimeric or fully-human),
antibody fragments, and the like. Often polynucleotides encoding
the isolated CDRs will be grafted into polynucleotides encoding
a suitable variable region framework (and optionally constant
regions) to form polynucleotides encoding complete antibodies

20   (e.g., humanized or fully-human), antibody fragments, and the
like. Antibodies can be used to isolate preparative quantities
of the antigen by immunoaffinity chromatography. Various other
uses of such antibodies are to diagnose and/or stage disease
(e.g., neoplasia), and for therapeutic application to treat

25   disease, such as for example: neoplasia, autoimmune disease,
AIDS, cardiovascular disease, infections, and the like.

        Various methods have been reported for increasing the
combinatorial diversity of a scFv library to broaden the
repertoire of binding species (idiotype spectrum). The use of

30   PCR has permitted the variable regions to be rapidly cloned
either from a specific hybridoma source or as a gene library from
non-immunized cells, affording combinatorial diversity in the
assortment of V_H and V_L cassettes which can be combined.
Furthermore, the V_H and V_L cassettes can themselves be

35   diversified, such as by random, pseudorandom, or directed
mutagenesis. Typically, V_H and V_L cassettes are diversified in
or near the complementarity-determining regions (CDRs), often the
third CDR, CDR3. Enzymatic inverse PCR mutagen sis has been

shown to be a simple and reliable method for constructing
relatively large libraries of scFv site-directed mutants (Stemmer
et al. (1993) _Biotechniques_ _14_: 256), as has error-prone PCR and
chemical mutagenesis (Deng et al. (1994) _J. Biol. Chem._ _269_:

5   9533). Riechmann et al. (1993) _Biochemistry_ _32_: 8848 showed
semirational design of an antibody scFv fragment using site-
directed randomization by degenerate oligonucleotide PCR and
subsequent phage display of the resultant scFv mutants. Barbas
et al. (1992) _op.cit._ attempted to circumvent the problem of

10   limited repertoire sizes resulting from using biased variable
region sequences by randomizing the sequence in a synthetic CDR
region of a human tetanus toxoid-binding Fab.

       CDR randomization has the potential to create
approximately 1×10^{20} CDRs for the heavy chain CDR3 alone, and

15   a roughly similar number of variants of the heavy chain CDR1 and
CDR2, and light chain CDR1-3 variants. Taken individually or
together, the combinatorics of CDR randomization of heavy and/or
light chains requires generating a prohibitive number of
bacteriophage clones to produce a clone library representing all

20   possible combinations, the vast majority of which will be non-
binding. Generation of such large numbers of primary
transformants is not feasible with current transformation
technology and bacteriophage display systems. For example,
Barbas et al. (1992) _op.cit._ only generated 5×10^7 transformants,

25   which represents only a tiny fraction of the potential diversity
of a library of thoroughly randomized CDRs.

       Despite these substantial limitations, bacteriophage
display of scFv have already yielded a variety of useful
antibodies and antibody fusion proteins. A bispecific single

30   chain antibody has been shown to mediate efficient tumor cell
lysis (Gruber et al. (1994) _J. Immunol._ _152_: 5368).
Intracellular expression of an anti-Rev scFv has been shown to
inhibit HIV-1 virus replication _in vitro_ (Duan et al. (1994)
Proc. Natl. Acad. Sci. (USA) _91_: 5075), and intracellular

35   expression of an anti-p21^{ras} scFv has been shown to inhibit
meiotic maturation of _Xenopus_ oocytes (Biocca et al. (1993)
Biochem. Biophys. Res. Commun. _197_: 422. Recombinant scFv which
can be used to diagnose HIV infection have also been reported,

demonstrating the diagnostic utility of scFv (Lilley et al.
(1994) <u>J. Immunol. Meth.</u> <u>171</u>: 211). Fusion proteins wherein an
scFv is linked to a second polypeptide, such as a toxin or
fibrinolytic activator protein, have also been reported (Holvost
5   et al. (1992) <u>Eur. J. Biochem.</u> <u>210</u>: 945; Nicholls et al. (1993)
J. Biol. Chem. <u>268</u>: 5302).

         If it were possible to generate scFv libraries having
broader antibody diversity and overcoming many of the limitations
of conventional CDR mutagenesis and randomization methods which
10   can cover only a very tiny fraction of the potential sequence
combinations, the number and quality of scFv antibodies suitable
for therapeutic and diagnostic use could be vastly improved. To
address this, the <u>in vitro</u> and <u>in vivo</u> shuffling methods of the
invention are used to recombine CDRs which have been obtained
15   (typically via PCR amplification or cloning) from nucleic acids
obtained from selected displayed antibodies. Such displayed
antibodies can be displayed on cells, on bacteriophage particles,
on polysomes, or any suitable antibody display system wherein the
antibody is associated with its encoding nucleic acid(s). In a
20   variation, the CDRs are initially obtained from mRNA (or cDNA)
from antibody-producing cells (e.g., plasma cells/splenocytes
from an immunized wild-type mouse, a human, or a transgenic mouse
capable of making a human antibody as in WO92/03918, WO93/12227,
and WO94/25585), including hybridomas derived therefrom.

25   Polynucleotide sequences selected in a first selection
round (typically by affinity selection for displayed antibody
binding to an antigen (e.g., a ligand) by any of these methods
are pooled and the pool(s) is/are shuffled by <u>in vitro</u> and/or <u>in
vivo</u> recombination, especially shuffling of CDRs (typically
30   shuffling heavy chain CDRs with other heavy chain CDRs and light
chain CDRs with other light chain CDRs) to produce a shuffled
pool comprising a population of recombined selected
polynucleotide sequences. The recombined selected polynucleotide
sequences are expressed in a selection format as a displayed
35   antibody and subjected to at least one subsequent selection
round. The polynucleotide sequences selected in the subsequent
selection round(s) can be used directly, sequenced, and/or
subjected to one or more additional rounds of shuffling and

subsequent selection until an antibody of the desired binding
affinity is obtained. Selected sequences can also be backcrossed
with polynucleotide sequences encoding neutral antibody framework
sequences (i.e., having insubstantial functional effect on
5   antigen binding), such as for example by backcrossing with a
human variable region framework to produce human-like sequence
antibodies. Generally, during backcrossing subsequent selection
is applied to retain the property of binding to the predetermined
antigen.

10   Alternatively, or in combination with the noted
variations, the valency of the target epitope may be varied to
control the average binding affinity of selected scFv library
members. The target epitope can be bound to a surface or
substrate at varying densities, such as by including a competitor
15   epitope, by dilution, or by other method known to those in the
art. A high density (valency) of predetermined epitope can be
used to enrich for scFv library members which have relatively low
affinity, whereas a low density (valency) can preferentially
enrich for higher affinity scFv library members.

20   For generating diverse variable segments, a collection
of synthetic oligonucleotides encoding random, pseudorandom, or
a defined sequence kernal set of peptide sequences can be
inserted by ligation into a predetermined site (e.g., a CDR).
Similarly, the sequence diversity of one or more CDRs of the
25   single-chain antibody cassette(s) can be expanded by mutating the
CDR(s) with site-directed mutagenesis, CDR-replacement, and the
like. The resultant DNA molecules can be propagated in a host
for cloning and amplification prior to shuffling, or can be used
directly (i.e., may avoid loss of diversity which may occur upon
30   propagation in a host cell) and the selected library members
subsequently shuffled.

                Displayed peptide/polynucleotide complexes (library
members) which encode a variable segment peptide sequence of
interest or a single-chain antibody of interest are selected from
35   the library by an affinity enrichment technique. This is
accomplished by means of a immobilized macromolecule or epitope
specific for the peptide sequence of interest, such as a
receptor, other macromolecule, or other epitope species.

Repeating the affinity selection procedure provides an enrichment
of library members encoding the desired sequences, which may then
be isolated for pooling and shuffling, for sequencing, and/or for
further propagation and affinity enrichment.

5   The library members without the desired specificity are
removed by washing.   The degree and stringency of washing
required will be determined for each peptide sequence or single-
chain antibody of interest and the immobilized predetermined
macromolecule or epitope.   A certain degree of control can be
10   exerted over the binding characteristics of the nascent
peptide/DNA complexes recovered by adjusting the conditions of
the binding incubation and the subsequent washing.   The
temperature, pH, ionic strength, divalent cations concentration,
and the volume and duration of the washing will select for
15   nascent peptide/DNA complexes within particular ranges of
affinity for the immobilized macromolecule. Selection based on
slow dissociation rate, which is usually predictive of high
affinity, is often the most practical route.   This may be done
either by continued incubation in the presence of a saturating
20   amount of free predetermined macromolecule, or by increasing the
volume, number, and length of the washes.   In each case, the
rebinding of dissociated nascent peptide/DNA or peptide/RNA
complex is prevented, and with increasing time, nascent
peptide/DNA or peptide/RNA complexes of higher and higher
25   affinity are recovered.

          Additional modifications of the binding and washing
procedures may be applied to find peptides with special
characteristics. The affinities of some peptides are dependent
on ionic strength or cation concentration. This is a useful
30   characteristic for peptides that will be used in affinity
purification of various proteins when gentle conditions for
removing the protein from the peptides are required.

          One variation involves the use of multiple binding
targets (multiple epitope species, multiple receptor species),
35   such that a scFv library can be simultaneously screened for a
multiplicity of scFv which have different binding specificities.
Given that the size of a scFv library often limits the diversity
of potential scFv sequences, it is typically desirable to us scFv

libraries of as large a size as possible. The time and economic considerations of generating a number of very large polysome scFv-display libraries can become prohibitive. To avoid this substantial problem, multiple predetermined epitope species

5   (receptor species) can be concomitantly screened in a single library, or sequential screening against a number of epitope species can be used. In one variation, multiple target epitope species, each encoded on a separate bead (or subset of beads), can be mixed and incubated with a polysome-display scFv library

10   under suitable binding conditions. The collection of beads, comprising multiple epitope species, can then be used to isolate, by affinity selection, scFv library members. Generally, subsequent affinity screening rounds can include the same mixture of beads, subsets thereof, or beads containing only one or two

15   individual epitope species. This approach affords efficient screening, and is compatible with laboratory automation, batch processing, and high throughput screening methods.

A variety of techniques can be used in the present invention to diversify a peptide library or single-chain antibody

20   library, or to diversify, prior to or concomitant with shuffling, around variable segment peptides or V_H, V_L, or CDRs found in early rounds of panning to have sufficient binding activity to the predetermined macromolecule or epitope. In one approach, the positive selected peptide/polynucleotide complexes (those

25   identified in an early round of affinity enrichment) are sequenced to determine the identity of the active peptides. Oligonucleotides are then synthesized based on these active peptide sequences, employing a low level of all bases incorporated at each step to produce slight variations of the

30   primary oligonucleotide sequences. This mixture of (slightly) degenerate oligonucleotides is then cloned into the variable segment sequences at the appropriate locations. This method produces systematic, controlled variations of the starting peptide sequences, which can then be shuffled. It requires,

35   however, that individual positive nascent peptide/polynucleotide complexes be sequenced before mutagenesis, and thus is useful for expanding the diversity of small numbers of recovered complexes and selecting variants having higher binding affinity and/or

DORSEY & WHITNEY LLP

MINNEAPOLIS
SEATTLE
NEW YORK
WASHINGTON, D.C.
DENVER
LONDON
SAN FRANCISCO
ANCHORAGE
DES MOINES
SALT LAKE CITY
SOUTHERN CALIFORNIA
FARGO

FOUR EMBARCADERO CENTER
SUITE 3400
SAN FRANCISCO, CALIFORNIA 94111-4187
TELEPHONE: (415) 781-1989
FAX: (415) 398-3249
www.dorseylaw.com

David H. Rogers, Ph.D.
Technical Advisor
rogers.dave@dorseylaw.com

BRUSSELS
TOKYO
PALO ALTO
GREAT FALLS
ROCHESTER
TORONTO
HONG KONG
MISSOULA
NORTHERN VIRGINIA
SHANGHAI
VANCOUVER

26 September 2002

James J. Diehl, Ph.D.
Senior Patent Counsel
RIGEL PHARMACEUTICALS, INC.
240 East Grand Avenue
South San Francisco, CA 94080

Re: European Application No. 00979200.3-2405
        Applicant: RIGEL PHARMACEUTICALS, INC.
        Filed: 17 November 2000
        For: **ING2, IAPS ASSOCIATED CELL CYCLE**
                      PROTEIN, COMPOSITIONS ND METHODS OF USE
        Our File No.: FE-68289-1-EP/RMS/DHR
        Dorsey & Whitney Docket No: 467802-344

Dear Jim:

    Enclosed for your information and files is a copy of a letter from our European associates dated 11 September 2002, together with a copy of the amended claims, confirming that they have filed the amended claims in accordance with our instructions.

    I am working with Robin Silva on this matter. If you have any questions or concerns, please do not hesitate to contact us.

Sincerely,

DORSEY & WHITNEY LLP

David H. Rogers, Ph.D.

DHR/mk
Enclosure

cc: Dolly A. Vance, Esq., RIGEL PHARMACEUTICALS, INC. (w/o encl.)
        Ms. Lori Sandoval, RIGEL PHARMACEUTICALS, INC. (w/o encl.)
        Robin M. Silva, Esq. (w/o encl.)

SF-1093957_1